

ТРЕБОВАНИЯ, ПРЕДЪЯВЛЯЕМЫЕ К ПЕРСПЕКТИВНОЙ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЕ АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА

Анализ тональности текста является одной из наиболее перспективных задач в области автоматической обработки текста (Natural Language Processing). Большинство работ в данной области появились за последние 10-12 лет, что связано с развитием социальных сетей, блог-платформ и других информационных технологий. Контент-анализ субъективности необходим для успешного функционирования вопросно-ответных систем и систем интеллектуальной обработки информации, где требуется отличать факты от мнений. Кроме того, производители товаров и поставщики услуг заинтересованы в получении обработанной информации о настроениях потребителей. Потребителей, в свою очередь, при выборе товара или услуги, интересуют мнения других людей, основанные на их личном опыте.

При автоматическом определении тональности текста выделяют следующие направления:

1) на основе правил с использованием шаблонов (rule-based with patterns). Данный метод заключается в генерации правил, на основе которых определяется тональность текста. Для этого текст разбивается на слова или последовательности слов (N-grams). Затем полученные данные используются для выделения часто встречающихся шаблонов, которым присваивается положительная или отрицательная оценка. Выделенные шаблоны применяются при создании правил вида «ЕСЛИ условие, ТО заключение»;

2) машинное обучение без учителя (unsupervised learning). Данный метод основан на идее, что наибольший вес в тексте имеют термины, которые чаще встречаются в этом тексте и в то же время присутствуют в небольшом количестве текстов всей коллекции. Выделяются данные тер-

мины и определяется их тональность. На основании этого делается вывод о тональности всего текста;

3) машинное обучение с учителем (supervised learning). В этом методе требуется наличие обучающей коллекции размеченных в рамках эмотивного пространства текстов, на базе которой строится статистический или вероятностный классификатор;

4) гибридный метод (hybrid method). Данный подход сочетает все или любые два из рассмотренных выше методов и заключается в применении классификаторов на их основе в определенной последовательности.

Анализ методов, положенных в основу работы существующих систем автоматического определения тональности текста, позволил выработать основные требования к перспективной интеллектуальной системе анализа тональности текста (ИСАТТ):

- в системе должна учитываться специфика русского языка – его морфология, свободный порядок слов, семантика текста и т.д. – что напрямую влияет на эффективность работы ИСАТТ;

- оценка тональности текста должна производиться по более широкой шкале, чем бинарная;

- ИСАТТ должна не только определять факт наличия эмоциональной окраски, но и вычислять его силу;

- для определения силы эмоциональной окраски текста возможно использование средств фоносемантической оценки слов и предложений;

- результат работы ИСАТТ должен выдаваться в простой и понятной форме, доступной к использованию не специалистами.

На сегодняшний день ни одна из разработанных ИСАТТ не может обеспечить пользователя качественной оценкой тональности текста. А ведь количество неструктурированной информации, представленной в текстовом виде, постоянно растет. Поэтому задача разработки математического и лингвистического обеспечения интеллектуальных систем анализа тональности текста является не просто актуальной, но и необходимой.